Military Operations Research Society

73rd MORSS CD Cover Page

712CD

For office use only 41205

UNCLASSIFIED DISCLOSURE FORM CD Presentation

21-23 June 2005, at US Military Academy, West Point, NY

Please complete this form 712CD as your cover page to your electronic briefing submission to the MORSS CD. Do not fax to the MORS office.

Author Request (To be completed by applicant) - The following author(s) request authority to disclose the following presentation in the MORSS Final Report, for inclusion on the MORSS CD and/or posting on the MORS web site.

Name of Principal Author and all other author(s):
Steve Wilcox, Ph.D.
Pete Bulanow

Principal Author's Organization and address:
Northrop Grumman Information Technology
2100 Washington Boulevard
Arlington, VA 22204

Phone: 703-312-2511

Fax: 703-312-2780

Email: steve.wilcox@ngc.com

Original title on 712 A/B: Evaluating Statistical Assumptions When Analyzing Deterministic Simulation Results Using a Design of Experiments

Revised title:____

Presented in (input and Bold one): (**WG 26 & 31**, CG___, Special Session ___, Poster, Demo, or Tutorial)

This presentation is believed to be:
UNCLASSIFIED AND APPROVED FOR PUBLIC RELEASE

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
22 JUN 2005	**N/A**	**-**

4. TITLE AND SUBTITLE
Evaluating Statistical Assumptions When Analyzing Deterministic Simulation Results Using a Design of ExperimentsEvaluating Statistical Assumptions When Analyzing Determinist

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Northrop Grumman Information Technology 2100 Washington Boulevard Arlington, VA 22204

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

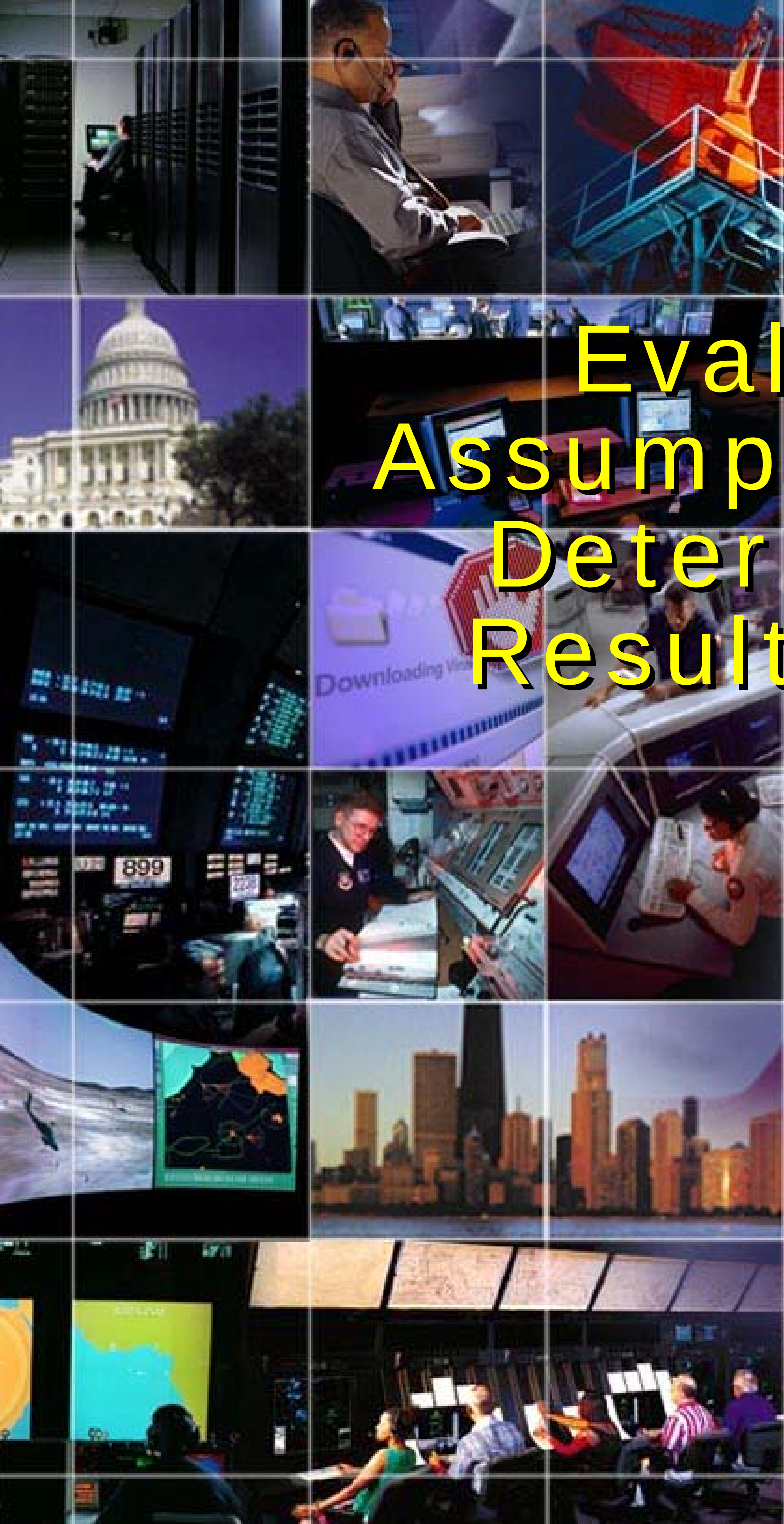
12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM201946, Military Operations Research Society Symposium (73rd) Held in West Point, NY on 21-23 June 2005. , The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**18**	



NORTHROP GRUMMAN

Evaluating Statistical Assumptions When Analyzing Deterministic Simulation Results Using a Design of Experiments

. Steve Wilcox, Ph.D.
Pete Bulanow

Northrop Grumman
Information Technology Sector

June 2005

Overview

- Background
  - Value Added Analysis process using Vector-In-Commander (VIC) to assess trades
  - Standard practice of employing fractional factorial experimental designs plus Analysis of Variance (ANOVA)
  - Non-monotonic relationships between parameters and VIC outcomes
- Issue
  - Statistical assumptions underlying ANOVA in this context
- Questions
  - Are these assumptions valid?
  - How to test them and assess impact?
- Analysis methodology
  - Graphical evidence
  - Independent replications of a small study whose parameters are known to have an expected value of zero
- Finding
  - Evidence that the ANOVA assumptions are not met
  - P-values from ANOVA are understated

The Value Added Analysis Process

- Supporting the Center for Army Analysis (CAA)
- Uses the Vector-in-Commander (VIC) Corps-level combat simulation model
- Objective:
  - Estimate the incremental contribution of system trades to combat effectiveness
  - Perform a cost-benefit analysis to determine the actual 'value-added' of the systems of interest.

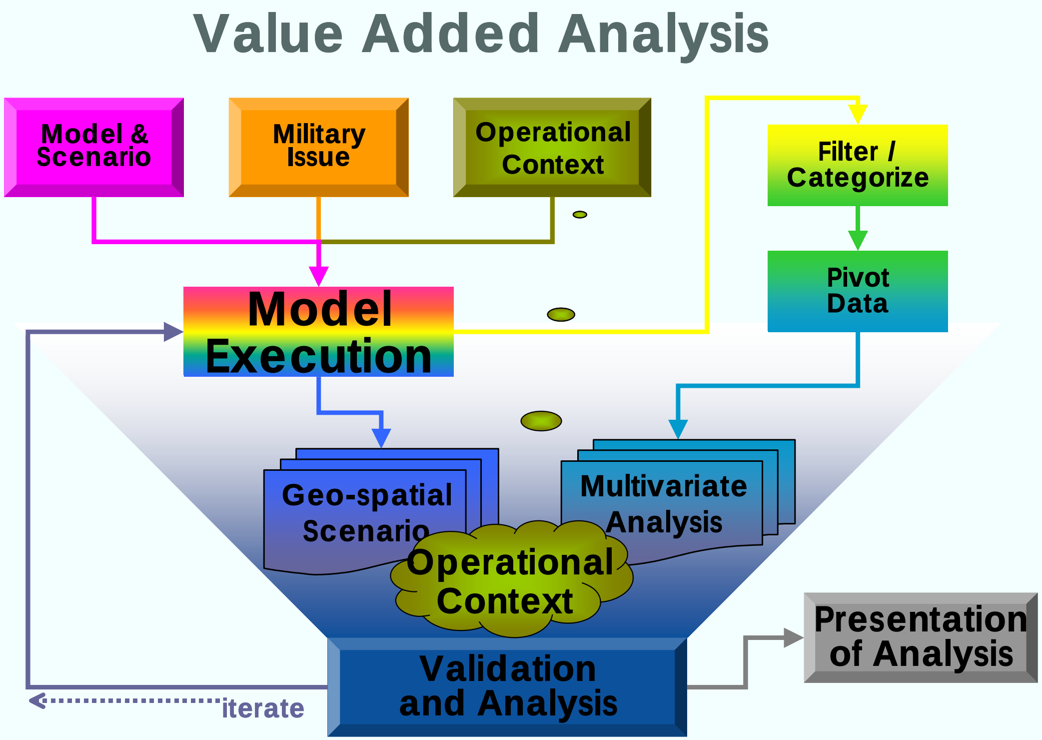


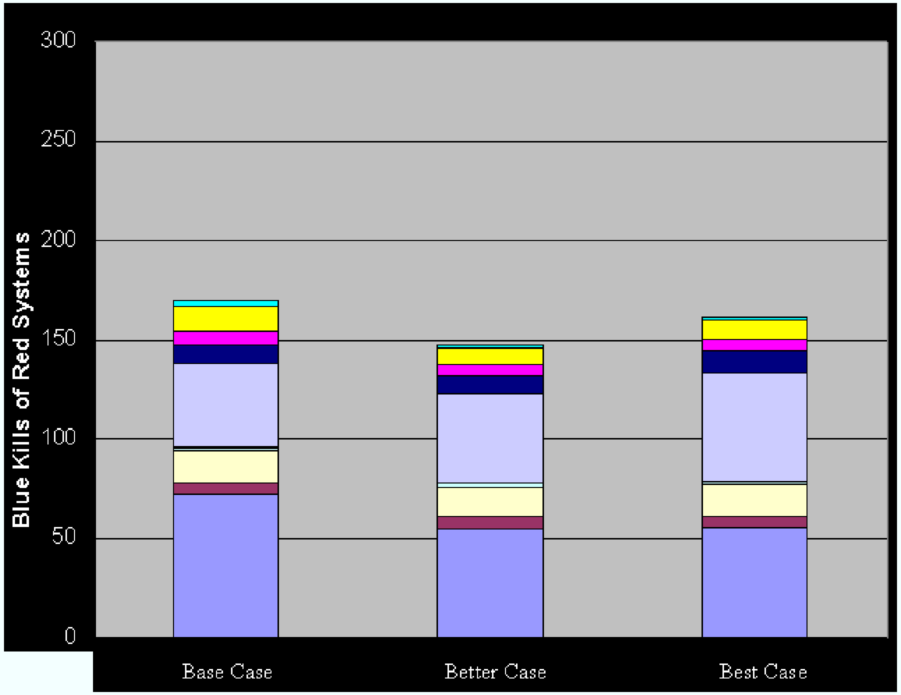
Figure 1. The VAA Process

- Previous methodology was a typical Design of Experiments (DOE) approach
- Now a perturbation methodology induces <u>stochastic</u> behavior in VIC

Non-Monotonicity in Deterministic Combat Models

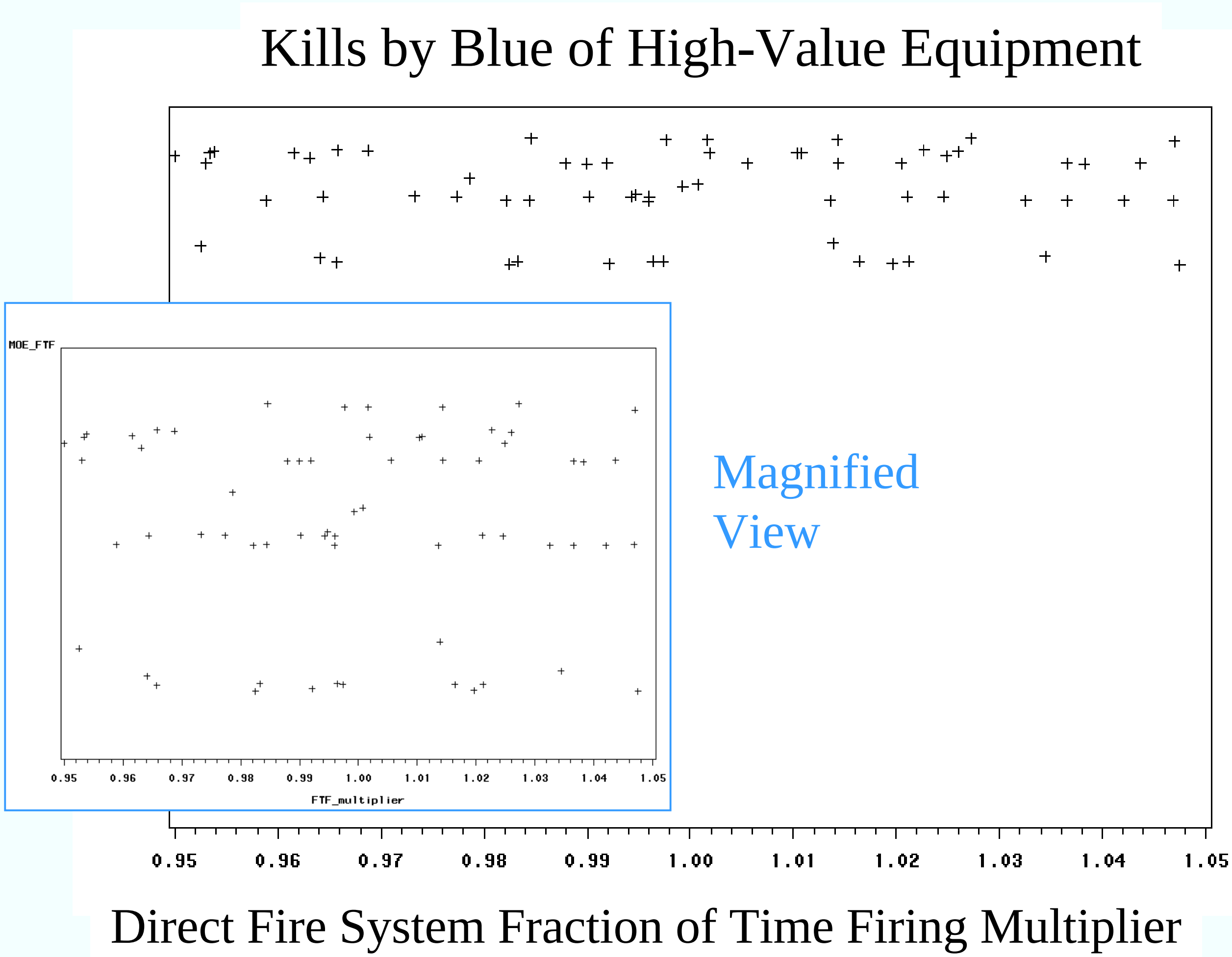
- A.K.A. Sensitivity to initial conditions and parameter values
  - Extensively noted in toy models of combat
    - The RAND model (Dewar, et al, 1991)
  - Also noted in VIC
    - Saeger & Hinch (2001)
    - Geoff Hawkins (1984) with VECTOR-2
- Impedes analysis of equipment trades

Kills of Selected Equipment by a Blue System of Interest



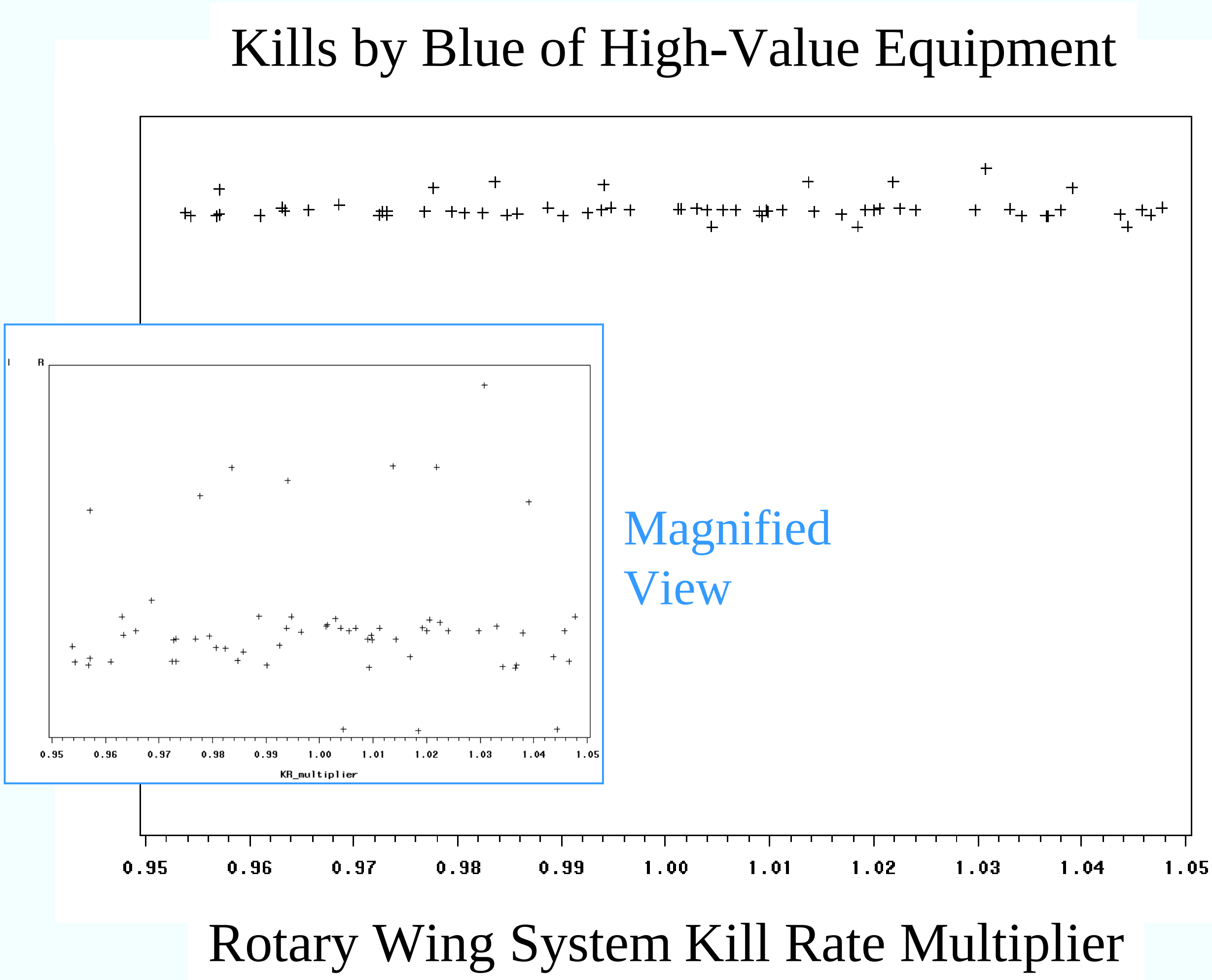
Parametric Sensitivity

- **The direct fire system (DF Sys) fraction of time firing (FTF) is multiplied by a number randomly selected from the interval (0.95, 1.05)**
- **Blue kills vary non-monotonically**
  - **and significantly**
- **If the analyst had chosen a very slightly different number, the MOE would be different**



Another Example

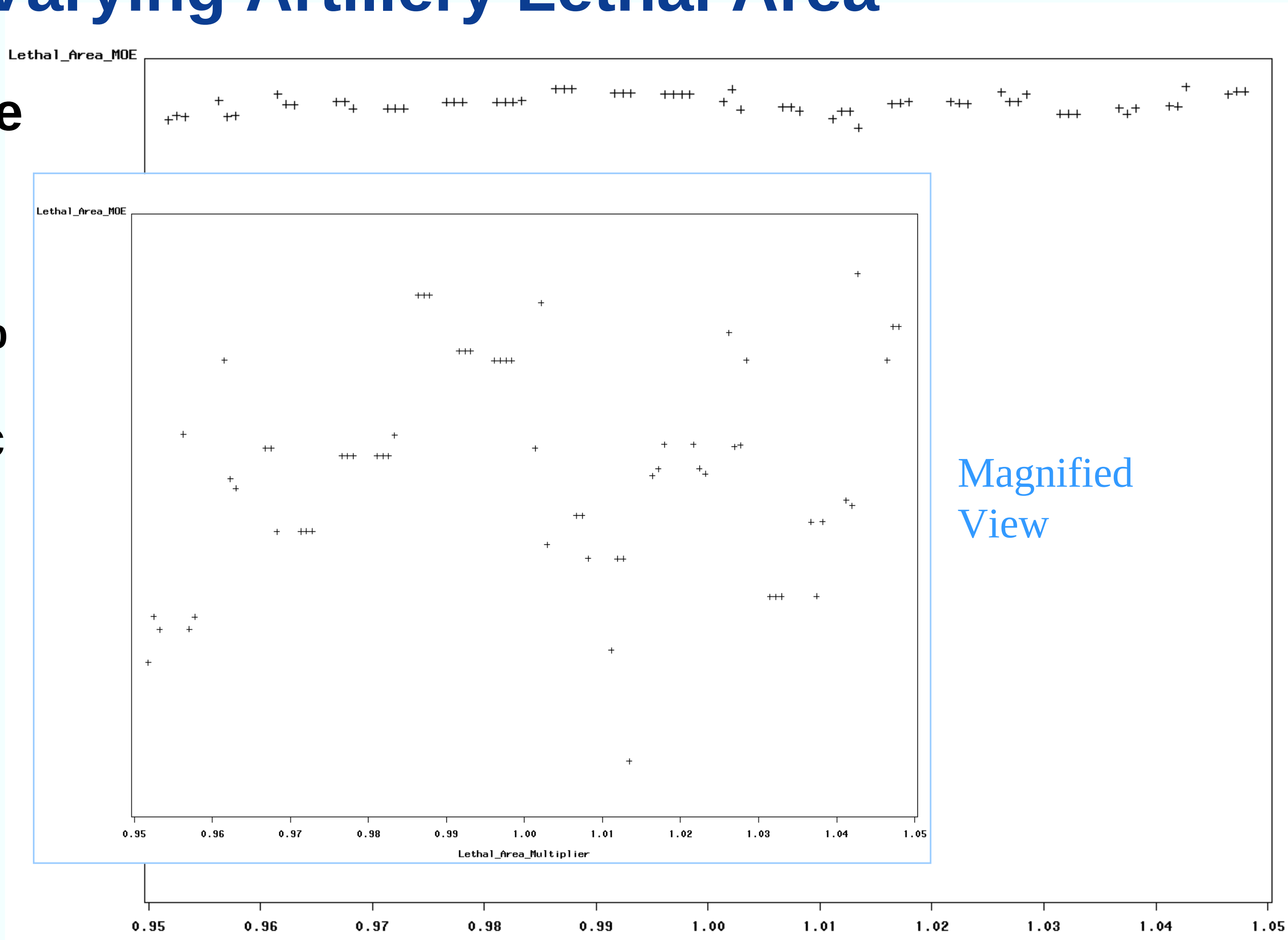
- **The rotary wing system (RW Sys) “kill rate” (KR) is multiplied by a number randomly selected from the interval (0.95, 1.05)**
- **Over this small interval, the true relationship is relatively weak**
  - **Non-systematic variation is much more important than the systematic relationship**



NORTHROP GRUMMAN

Slightly Varying Artillery Lethal Area

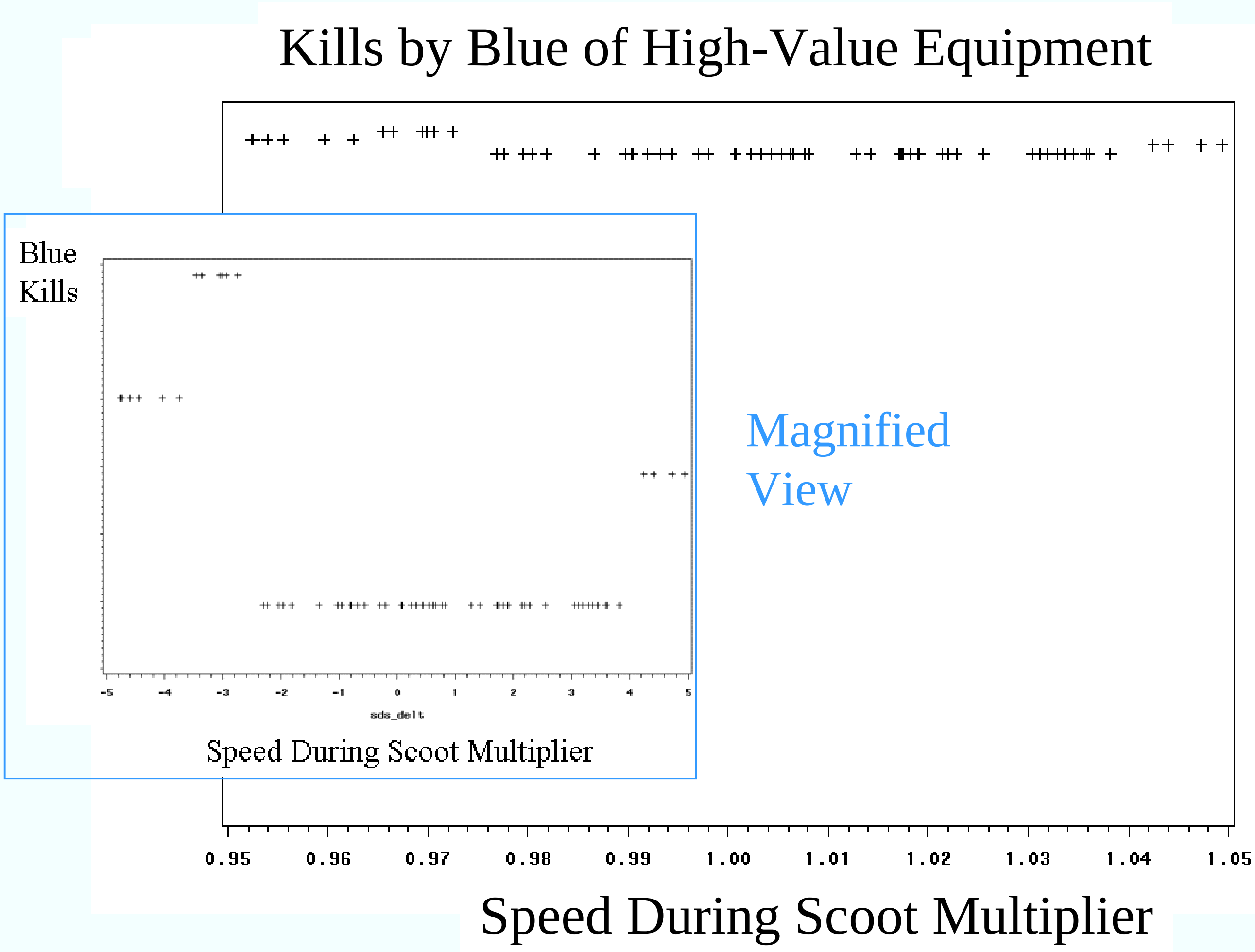
- Varying the artillery munition lethal area within ±5% has non-monotonic effects on the MOE
- Notice the flat spots



Lethal Area (LA) Multiplier

The Amount of Sensitivity is Not the Same for all Parameters

- **Varying speed during scoot ±5%**
- **Only four clusters appear in the multivariate data**
  - **When viewed in total or by victim equipment group**



Statistical Assumptions in the Design of Experiments

- **A typical DOE is to run a number of combinations of experimental settings**
  - **And then analyze the MOEs using analysis of variance**
- **Different parameter settings are superimposed for each run**
- **ANOVA makes the assumption that error terms are mean-zero and statistically independent**

Run	Sys1	Sys2	Sys3	Sys4	Sys5	Sys6	Sys7
1	1	1	1	1	1	1	1
2	1	1	1	-1	1	-1	-1
3	1	1	-1	1	-1	-1	-1
4	1	1	-1	-1	-1	1	1
5	1	-1	1	1	-1	-1	1
6	1	-1	1	-1	-1	1	-1
7	1	-1	-1	1	1	1	-1
8	1	-1	-1	-1	1	-1	1
9	-1	1	1	1	-1	1	-1
10	-1	1	1	-1	-1	-1	1
11	-1	1	-1	1	1	-1	1
12	-1	1	-1	-1	1	1	-1
13	-1	-1	1	1	1	-1	-1
14	-1	-1	1	-1	1	1	1
15	-1	-1	-1	1	-1	1	1
16	-1	-1	-1	-1	-1	-1	-1

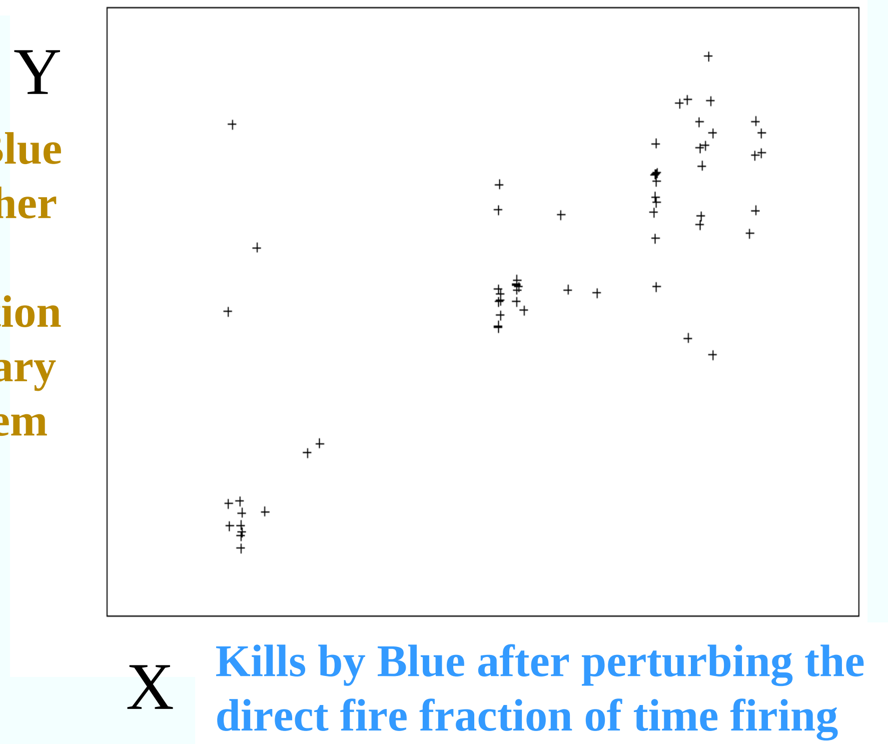
$$y_i = \alpha + \sum_{k=1}^{m} \beta_k x_{ki} + \varepsilon_i$$

$$E[\varepsilon_i \mid \mathbf{x}] = 0 \qquad i = 1, \ldots, n$$

$$Cov[\varepsilon_i, \varepsilon_k] = 0 \quad i = 1, \ldots, n; \quad k = 1, \ldots, m$$

Effect of Superimposing Parameter Settings

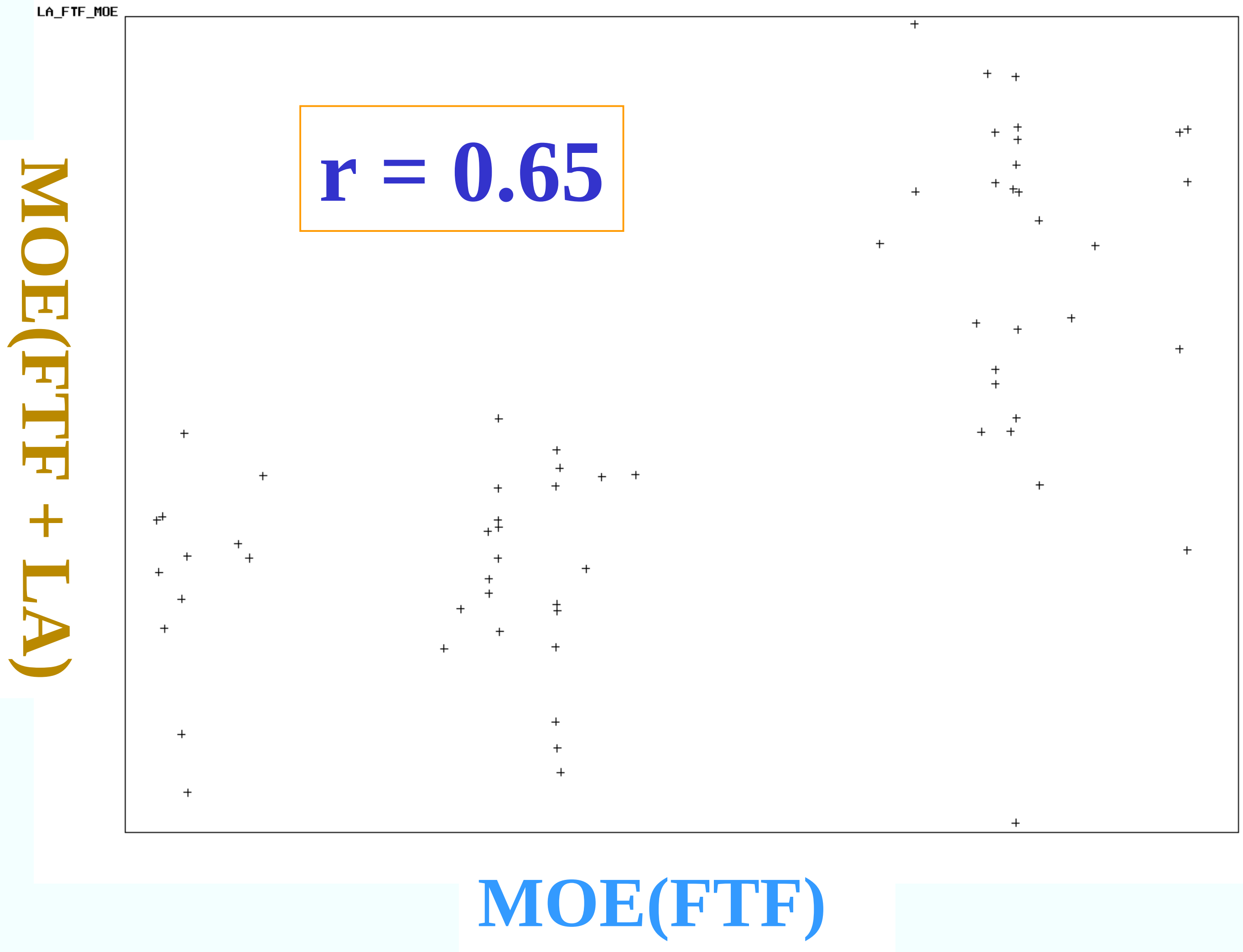
- A perturbation to the rotary wing system kill rate (KR) is superimposed on that of direct fire system fraction of time firing (FTF)
  1. Perturb FTF
  2. Perturb FTF and KR
- If the first run had a high MOE, the second one was also likely to be high
  - This is a “memory effect”
  - The artifacts from the first perturbation are “remembered” even though another perturbation is applied.



Run#	FTF	Kills(FTF)	KR	Kills(FTF+KR)
1	FTF#1	X1	KR#1	Y1
2	FTF#2	X2	KR#2	Y2
3	FTF#3	X3	KR#3	Y3
…	…	…	…	…
		(Seen Before)		(Combined Perturbation)

The "Memory Effect" As Correlation

- The variability of X and Y are essentially noise in the narrow parameter ranges employed.
- But the noise is "correlated" when a perturbation in one parameter is added to the perturbation of another



The Relationship Between DOEs and ANOVA Assumptions

An example DOE:

Run	FTF Option	KR Option
0	base	base
1	improved	base
2	base	improved
3	improved	improved

→

Run	x1	x2
0	0	0
1	1	0
2	0	1
3	1	1

When slightly different performance parameter values are chosen (as by different analysts or rounding), we have:

	Try #1		Try #2		Try #3		...
Run(i)	FTF Value	KR Value	FTF Value	KR Value	FTF Value	KR Value	
0	base	base	base	base	base	base	
1	FTF#1	base	FTF#2	base	FTF#3	base	
2	base	KR#1	base	KR#2	base	KR#3	
3	FTF#1	KR#1	FTF#2	KR#2	FTF#3	KR#3	

Ultimately the model being estimated is

$$MOE_i = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i$$

But ε_3 is correlated with ε_1 and ε_2 due to the memory effect.

This is a deviation from the ANOVA assumptions

A Model Study That Should Have No Findings

- **A small study with an expectation of zero findings is independently performed 64 times**
- **How often is "statistical significance" achieved?**
- **The true values of the parameters being estimated are zero**
  - **The deviations from the base case are small (±5%) and are centered on the base case values**
  - **The parameter settings are independently varied (as if a different analyst were assigned each time)**

Run Number	FTF	KR	LA
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	-1	1	1
5	1	-1	-1
6	1	-1	1
7	1	1	-1
8	1	1	1

More on the Model Study

- 64 independent repetitions of a 3-factor factorial design
- Factors
  - Fraction of time firing for a direct fire weapon system
  - “Kill rate” for a rotary wing system, and
  - An artillery munition lethal area
- MOE is kills by Blue’s high-value equipment of Red’s high-value equipment
- Analyzed using a 3-factor main effects unbalanced ANOVA (least squares regression) model
  - One observation is missing in a number of repetitions
  - More than one error d.f. for robustness of findings
- ANOVA results are analyzed statistically

Results For the ANOVA Overall F Statistic

Dependent Variable	Average P-Value	P-Value of F (mean=0.5)	Number Statistically Significant	t statistic (versus 5% level of significance)
Blue Kills Red Important Equipment	0.28	2.31E-09	9	2.069*
Red Kills Blue Important Equipment	0.27	1.00E-09	15	3.454**

- The chance of Type I error exceeds the alpha of 5%
- P-Values from ANOVA will not actually have the nominal statistical significance level
- Similar results were found for individual ANOVA model parameters

Conclusions and Way Forward

- P-values from ANOVA applied to DOE data from deterministic combat simulations may not mean what one expects
  - Due to the “memory effect” seen in VIC
- The way forward is to employ perturbation to randomize the error terms in the ANOVA model
  - See Bulanow, et al 2004 MORSS presentation.
  - This erases the “memory effect” and assures error term independence for all practical purposes.

References

- Randomization of VIC
  - Pete Bulanow, Steve Wilcox, Angela Hunter, Michelle Enwright, Chris Hill, Gregory Heck, "Development of Randomization Methodologies for Analysis Using VIC", June 24, 2004, Military Operations Research Society Symposium
  - Bailey, Timothy (2001) Multiple Runs of a Deterministic Combat Simulation, TRADOC Analysis Center report TRAC-F-TD-02-001, November 2001.
  - Snare, Ross (2002) An Alternative Approach for Conducting Analysis with Deterministic Combat Simulations, Ph.D. dissertation, Walden University.
- Sensitivity of VIC Settings
  - Hawkins, G. (1984) "Structural Variance and Other Related Topics Experienced in the SHAPE Armour/Anti-Armour Study", pp. 627-636 in Reiner K. Huber, Systems Analysis and Modeling in Defense: Development, Trends, and Issues, New York: Plenum Press, 1984.
  - Saeger, K.J. and J.H. Hinch (2001) "Understanding Instability in a Complex Deterministic Combat Simulation", Military Operations Research, 6 (4), 43-55.
- Non-Monotonicity in General
  - Dewar, J.A., Gillogly, J.J., and M.L. Juncosa (1991) *Non-Monotonicity, Chaos, and Combat Models*, RAND R-3995-RC.
- Analytical Software
  - SAS Institute, Inc. (1999) SAS OnlineDoc®, Version 8, Cary, NC: SAS Institute, Inc.
- Design of Experiments
  - Douglas C. Montgomery (2000) Design and Analysis of Experiments, 5th Edition, Wiley.